

## 《心理学报》审稿意见与作者回应

题目：基于问题解决式的类比推理的老年：表面相似性和结构相似性的作用

作者：李美佳 庄丹琪 彭华茂

### 第一轮

审稿人 1 意见：

意见 1：被试选取方面

Q1 进一步明确青年人群和老年人群的入组标准：招募多少人，筛除多少人，剔除标准是什么？关于老年被试，请明确“无脑部疾病或神经退行性疾病”的具体疾病类型。

回应：感谢您的建议，本研究共招募 29 名年轻人和 32 名健康老年人，无脑部疾病或神经退行性疾病，如阿尔兹海默症。所有被试视力或矫正视力正常，由于研究是在一对一访谈的情况下进行的，故在保证被试正确理解任务的同时均完成任务，数据有效且完整，故而没有删减被试。同时，本研究测试了两年龄组被试的基本认知能力（言语理解、加工速度、工作记忆和抑制能力），与以往同类研究，选取的老年被试认知能力正常。

表 1 被试基本信息和背景测验汇总表（标准差）

	年轻人	老年人
年龄	21.86(1.432)	69.09(5.114)
受教育年限	14.48(1.022)	14.34(2.458)
词语理解得分	17.17(1.818)	16.65(2.496)
数字符号转换得分（个）	77.93(9.490)	44.97(9.687)
数字倒背广度（个）	8.24(1.327)	5.52(1.267)
Stroop 测验（个）	80.38(19.397)	48.78(17.270)

在正文部分（请见第 102 行，方法部分 2.1 被试），已对被试细节部分进行补充：

在\*大学招募 29 名 20~26 岁的年轻人（年龄： $M = 21.86$ ,  $SD = 1.43$ ；受教育年限  $M = 14.48$  年， $SD = 1.02$ ）；在\*市社区招募 32 名 62~79 岁的健康老年人（年龄： $M = 69.09$  岁， $SD = 5.11$ ；受教育年限： $M = 14.34$  年， $SD = 2.46$ ），无脑部疾病或神经退行性疾病，如阿尔兹海默症。所有被试视力或矫正视力正常，均完成任务，数据有效。两组被试在受教育年限 [ $M_{\text{年轻人}} = 14.48$ ,  $M_{\text{老年人}} = 14.34$ ;  $t(42) = 0.29$ ,  $p = 0.771$ ] 上无显著差异。类比推理任务涉及语义理解能力，故使用韦氏成人智力量表（城市版）中的解词测验对被试的语义理解能力进行测量，两组在词汇理解上亦无显著差异 [ $M_{\text{年轻人}} = 17.17$ ,  $M_{\text{老年人}} = 16.65$ ;  $t(59) = 0.94$ ,  $p = 0.353$ ]。实验后给予报酬。

**Q2** 完善被试的人口学资料。例如明确老年人群的年龄分布（最低和最高年龄值）；明确青年组和老年组的受教育背景等。

**回应：**感谢您的细心建议，老年人的年龄分布为 62~79 岁，受教育年限： $M = 14.34$  年；年轻人的年龄分布为 20~26 岁，受教育年限  $M = 14.48$  年。两部分均已在文中“**被试**”部分作出修改如下（请见第 102 行，方法部分 2.1 被试）：

在\*大学招募 29 名 20~26 岁的年轻人（年龄： $M = 21.86$ ,  $SD = 1.43$ ；受教育年限  $M = 14.48$  年， $SD = 1.02$ ）；在\*市社区招募 32 名 62~79 岁的健康老年人（年龄： $M = 69.09$  岁， $SD = 5.11$ ；受教育年限： $M = 14.34$  年， $SD = 2.46$ ），无脑部疾病或神经退行性疾病，如阿尔兹海默症。所有被试视力或矫正视力正常，均完成任务，数据有效。两组被试在受教育年限 [ $M_{\text{年轻人}} = 14.48$ ,  $M_{\text{老年人}} = 14.34$ ;  $t(42) = 0.29$ ,  $p = 0.771$ ] 上无显著差异。

**Q3** 实验前期选取了青年人和老年人来评估材料的有效性。这部分人群是否与被试组重合，请明确。问题解决范式的理论是知识的迁移，为保证试验结果的严谨性，应该保证材料评估人群和正式实验人群不重合。

**回应：**感谢您的提示，从科学严谨的角度来讲，材料评估应与参与正式实验的被试人群同质，但为了排除材料熟悉性带来的影响，材料评估的被试与正式实验亦不能重合。已在正文做出修改（请见第 142 行，方法部分 2.2 实验材料）：

## 2.2 实验材料

参与材料评估的被试与正式实验人群同质但不重合。

## 意见 2：文献引证

**Q1** 问题类比推理在认知水平和认知难度上都明显高于经典类比研究。对“问题解决范式更具生态效度”，可给出跟多的文献支撑以加强论据。文中原句：“基于问题情境的推理任务通常以故事或问题呈现任务，会比经典类比推理任务提供更多背景信息，可能会有助于老年人完成推理任务。”

**回应：**感谢您的建议，“问题解决范式更具生态效度”这句话确实更多来讲是一种推论，没有文献直接说明生态效度更高。从侧面来讲，格式塔心理学强调问题情境的结构的重要性，认为问题解决是形成问题情境的新的结构，即把握问题情境中诸事物的关系。故而，我们做出推论，在问题解决的情境中，或许会比经典的类比推理任务（A:B::C:D）提供更多的背景信息（context），可能有助于老年人完成推理任务。同时，以往经典的类比推理任务都是界定清晰的问题（well-defined problems），如河内塔，但现实生活中的问题界定往往不够清晰（ill-defined problems），这也是带来不同范式的结果差异的因素，后者更具生态性。此外，我们发现以往研究的结果中，基于类比推理的经典范式或问题解决范式上，存在研究结论的不一致，这或许也可以从侧面说明两种范式确实存在一些内部的差异，这些差异，如背景信息中包含的表面相似性或结构相似性，会不会成为帮助老年人提取源问题的关键呢？故而本研究关注了这一问题。

为了使得文章更为严谨，已在正文中对措辞进行修改：相关研究已补充至相应位置，以加强论据（请见第 38 行，引言部分）：

基于问题情境的推理任务通常以故事或问题呈现任务，需要被试根据长时记忆中得源问题将靶问题补充完整，需补充的部分不涉及故事主线（HOR），但模拟了真实的问题解决情境中的界定不清晰问题。相比于经典类比推理任务，基于问题情境的推理任务可提供更多背景信息（context），可能会有助于老年人完成推理任务。已有研究发现，在儿童研究中使用基于问题的类比推理任务，会比经典类比任务范式下表现的更好（Brown & Kane, 1988），通过视频呈现更多信息也会有助于推理（Chen & Siegler, 2013）。故而，本研究推论使用该范式将会比经典类比任务更有助于老年人完成推理任务。

同时，不同类型任务的详细说明在第一段也进行了完善（请见第 16 行，引言部分）：

类比推理研究中，推理任务形式包括两类。一类是皮亚杰提出、斯滕伯格（Sternberg & Nigro, 1980）改进的经典类比推理任务范式，其一般形式为  $a:b::c:d$ ，如面包：牛奶：：油条：豆浆，即人们可以由面包和牛奶的关系（西式早餐搭配）推测出油条和豆浆的关系（中式早餐搭配）。另一类是问题解决式的推理任务（Chen & Siegler, 2013），即面临新异问题（target）时，利用先前具有相似情形的问题解决方法（source）来解决当前问题，如被试阅读一个源问题：“樵夫李老头推一车柴下山，他口渴了便拿出水壶喝水。不料柴车着火，但周围没有水源，李老头情急之下赶紧把水壶中的水泼在柴车上，但由于水太少了，火势依旧凶猛。他只能眼看着自己的柴车被大火烧尽了”。在这个故事情境中，樵夫解决问题的方式是把身边可得到的液体（汤）泼洒在着火的车上，但只是杯水车薪，无法熄灭大火。之后被试阅读靶问题——“农民老张摘了一车的草准备运回家喂牲口。老张有些饿了，就停下来休息，拿出一碗汤。不料天气太热，老张发现草正在被烤焦，发出刺鼻的味道”——并回答此故事情境中接下来会发生什么事情。根据源问题中类似的情节发展，应当是“老张情急之下赶紧把手中的汤洒在草上，但只是杯水车薪，老张只能眼看着车里的草全被烤干了”。

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology*, 20(4), 493-523.

Chen, Z., & Siegler, R. S. (2013). Young children's analogical problem solving: Gaining insights from video displays. *Journal of Experimental Child Psychology*, 116(4), 904-913.

本研究的结果中，确实在老年人中发现了范式不同带来的差异（请见第 249 行，讨论部分）：

本研究发现的基于问题解决式的类比推理成绩的年龄差异，相比于经典类比推理任务的年龄差异较小。对比两种范式，以年轻人推理成绩的标准差为标准，用两年龄组的各自推理得分之差除以该范式下年轻人标准差（Gerstorff, Ram, Hoppmann, Willis, & Schaie, 2011），以此衡量不同任务范式下推理成绩的年龄差异。可以发现，在经典类比推理范式下，老年人的成绩比年轻人低 3.20SD（老年人：1.08 $\pm$ 0.87；年轻人：6.82 $\pm$ 1.79；高悦，彭华茂，王大华和文静，2014）或 1.73SD（老年人：59.9 $\pm$ 9.2；年轻人：79.1 $\pm$ 11.1；Salthouse, 1992b）；而在本研究中老年人的成绩比年轻人低 0.86SD（老年人：0.38 $\pm$ 0.25；年轻人：0.62 $\pm$ 0.28）。由此可见，在使用更具生态效度的问题解决范式时，老年人的推理成绩有所提升，与年轻人差距更小。总的来说，这种比较方法源于追踪研究中对不同同辈群体（cohort）的比较（Hülür, Ram, & Gerstorff, 2015; Schilling, 2005），但跨研究数据之间的比较是不够严谨的，只能作为粗略探讨。

**Q2** 文献显示类比推理能力表现出老化特征，是否可以列举更详细的文献来说明具体表现有哪些。

**回应：**感谢您的建议，以往研究通过老年人-年轻人的对比研究，或通过研究老年人的大脑白质完整性等，对推理老化的机制进行了解释说明；实验材料多样，以经典类比推理范式为主，如选用智力测验或成套认知能力测验中的词语推理（Bugaiska & Thibaut, 2015；Clark, Gardner, Brown, & Howell, 1990）、字母序列归纳推理（Saczynski, Willis, & Schaie, 2002）、已知前提类比句子关系（Salthouse, Legg, Palmon, & Mitchell, 1990）等。较少有研究关注问题解决中老年人的表现。

文章第三段已有补充说明（请见第 29 行，引言部分）：

以往的研究通过词语推理（Bugaiska & Thibaut, 2015; Clark, Gardner, Brown, & Howell, 1990）、字母序列归纳推理（Saczynski, Willis, & Schaie, 2002）、已知前提类比句子关系（Salthouse, Legg, Palmon, & Mitchell, 1990）等发现老年人的类比推理能力出现普遍的下降，老年人在类比推理中整合关系存在困难，难以抑制无关特征对正确信息的干扰作用（Viskontas, Holyoak, & Knowlton, 2005），以及较慢的反应时（Clark et al., 1990）等。在脑机制方面，研究者发现白质完整性（white matter integrity）与成套的认知能力测验中的推理能力及老年人的日常功能之间存在相关关系，前额叶皮层白质束和两半球间的交流在更高阶的认知功能中具有重要作用（Monge, Greenwood, Parasuraman, & Strenziok, 2016）。

**Q3** 类比推理（老化）领域，国内北京师范大学辛自强老师也有较多研究思考，可以参考阅读。

**回应：**辛老师基于已有理论和研究背景所提出了关系-表征复杂性模型，从分析问题中诸多集合间的关系入手刻画任务的复杂性，可以用作任务关系复杂性和儿童对其表征水平的分析框架，同时还研究了“基于关系-表征复杂性模型的数学应用题表征能力测验”等非常具有实用价值的课题，对于认知发展研究有较为普遍的意义。

该模型指出“关系复杂性”，包括关系的水平复杂性和等级复杂性，与本研究的两种相似性似乎具有某种映射关系：水平复杂性是简单累加性的，至少是等距变量，类似类比推理问题中的“表面相似性”；而等级复杂性是迭代式的增加，它是个等级变量且有传递性，类似本研究中的“结构相似性”。总的来说，研究者可对设计好的或选用的实验任务进行这种复杂性的分析，这种分析不需要依赖于被试的实际表现，而完全是研究者的一种逻辑的分析，旨在确定被试完成任务通常必须进行的最低层次的理解或者确定被试必须理解的任务中所包含的关系。但目前并未有文献对二者建立起联系，故我们不能妄自引用这种话语体系，但这种模型的思想值得借鉴，已将该文章作为问题提出的论据。

在文中修改部分如下（请见第 51 行，引言部分）：

发展研究中，对关系相似性的理解和表征是推理能力发展的重要内部基础（Gentner, 1988; Goswami, 1991; 辛自强, 2007），不论是经典推理任务（a:b::c:d）还是问题解决任务（Chen & Siegler, 2013）均如此。相似性会影响社会关系中对他人行为的表征

(Liviatan, Trope, & Liberman, 2008)、判断及决策偏差 (Koehler, Brenner, Liberman, & Tversky, 1996) 等。靶问题和源问题对应各成分的相似性可以促进对源问题的检索 (Catrambone, 2002; Holyoak, 1987)。

辛自强. (2007). 关系-表征复杂性模型. *心理发展与教育*, 3, 122-128.

Koehler, D. J., Brenner, L. A., Liberman, V., & Tversky, A. (1996). Confidence and accuracy in trait inference: Judgment by similarity. *Acta Psychologica*, 92(1), 33-57.

Liviatan, I., Trope, Y., & Liberman, N. (2008). Interpersonal similarity as a social distance dimension: Implications for perception of others' actions. *Journal of Experimental Social Psychology*, 44(5), 1256-1269.

.....

#### 审稿人 2 意见:

**意见:** 这篇文章思路清晰、结构和层次分明。问题把握精确、参考文献详实、较符合学术规范。文章的数据分析基本能够支撑文章所得的结论,有一定的创新性,是一篇较高质量的论文。编辑部可以考虑录用。

**回应:** 非常感谢您对本文的首肯和对研究的精确解读。祝您研究愉快!

.....

#### 审稿人 3 意见:

作者采用问题解决的实验范式考察了老年人类比推理与年轻人类比推理的差异及其受表面相似性与结构相似性的不同影响。研究比较有意义,有一定的创新,结论比较可靠,但存在如下问题:

**意见 1:** 研究的创新性贡献不是十分清晰。

**回应:** 非常感谢您的宝贵意见。结合以往研究,对本文的创新性贡献进行汇总如下:

#### ❖ 关注表面相似性和结构相似性哪个对老年人的推理作用更大是本研究的首要贡献;

表面相似性和结构相似性的作用谁更大一直存在争议 (similar debate)。如前人研究考察 5~6 岁, 9~10 岁和年轻成人类比推理的发展规律, 发现随着年龄增长, 被试越来越能够解释出结构相似, 并使用结构相似性解决靶问题 (Gentner, 1988; Markman, Gentner, 1996), 也即问题解决似乎更依赖于深层结构的相似性而非实体间的相似性。而新进的样例理论的研究结果 (Gilovich, 1981; Ross, 1984; Holyoak, 1987; Catrambone & Holyoak, 1989) 表明, 不仅是结构相似性, 突出的表面相似性也能够激活对源问题的检索, 从而促进类比推理。**总的来说, 前人研究中, 不同年龄的被试似乎存在不同的结论:** 结构相似性占优势的研究, 对象大多是年轻人 (Schnitz & Baadte, 2015; Holyoak & Koh, 1987); 而对于儿童和老年人, 则更优先加工表面相似性 (Viskontas et al., 2004)。以往有实验室研究认为, 老年人更多编码表面信息, 没有或较少地加工关系 (结构) 信息。此外, 在视觉工作记忆的研究中发现, 相比单独记忆颜色这一表面信息, 老年人将表面信息 (颜色-形状) 结合的能力出现衰退 (Isella, Molteni, Mapelli, & Ferrarese, 2015) 本研究关注在基于问题解决式的类比推理任务中, 老年人和年轻人对表面信息和结构信息相似性的敏感程度



是否存在差异？根据以往研究结果，相比结构相似，表面信息相似性的提高是否能更多促进老年人的推理？发展研究的不一致结果，**原因**可能是由于实验材料不够生态化、对比分析无法很好地剥离二者的交互带来的。在类比推理的阶段研究中，类比问题解决的一个关键在于能够从以往经验中提取出相应的源问题用于当前问题，但以往研究多关注编码阶段，老年人具有丰富的生活经验，似乎能更好地解决问题，但在日常生活中常发现类比失败，可推论，老年人可能是无法提取出对应的源问题中的信息。故而，**本研究**从毕生发展的视角探讨了相似性在类比推理中的作用，通过实验操纵使得老年人和年轻人都达到同等编码源问题的水平，探讨了信息相似性在类比推理中的作用，以及随增龄带来的影响。

❖ **关注老年人的基于问题解决式的类比推理能力是否出现老化。**

目前基于问题解决式的推理研究较少，且被试多为儿童（Tunteler, 2002）、青少年（Jablansky, Alexander, Dumas, & Compton, 2016）和大学生（Catrambone, 2002; Ross, 1989）。因此，老年人在基于问题解决的类比推理任务中，是否也表现出较差的成绩，是本研究考察的问题之一。

❖ **基于问题解决范式而非实验室范式，更具生态效度。**

以往的研究通过词语推理（Bugaiska & Thibaut, 2015）、字母序列归纳推理（Saczynski, Willis, & Schaie, 2002）、已知前提类比句子关系（Salthouse, Legg, Palmon, & Mitchell, 1990）老年人在实际生活中，有更多机会遇到可基于先前问题解决办法来应对对新问题的情况，且老年人自身的生活经验和智慧可能也会有助于老年人完成问题解决式的推理任务（Jeste & Oswald, 2014）

正文中，在引言和讨论部分进行了补充如下：

**（请见第 47 行，引言部分）**

目前基于问题解决式的推理研究较少，且被试多为儿童（Tunteler, 2002）、青少年（Jablansky, Alexander, Dumas, & Compton, 2016）和大学生（Catrambone, 2002; Ross, 1989）。因此，老年人在基于问题解决的类比推理任务中，是否也表现出较差的成绩，是本研究考察的问题之一。

**（请见第 65 行，引言部分）**

源问题和靶问题成分的相似性可促进检索源问题，从而提高推理成绩（Catrambone, 2002; Holyoak & Koh, 1987），但相似性作用复杂，表面相似性和结构相似性的作用谁更大一直存在争议（similar debate; Bulloch & Opfer, 2009）。

**（请见第 94 行，引言部分）**

综上所述，本研究基于问题解决范式，采用故事类比的任務，通过对源问题和靶问题中包含的表面信息和结构信息的相似性进行系统操纵，使得老年人和年轻人对源问题达到同等编码水平，旨在研究以下两个问题：（1）基于问题解决范式的类比推理，老年人的成绩相比于年轻人是否出现下降；（2）类比推理中表面相似和结构相似对不同年龄群体的推理质量是如何作用的：对于老年人而言，是否表面相似性更重要；对年轻人而言，是否结构相似性更重要？

**（请见第 246 行，讨论部分）**

本研究总体发现，在年龄特征上，老年人的类比推理成绩比年轻人差。源问题和靶问题

间表面相似性与结构相似性的提高会带来推理质量的提升。这与许多研究得出了相同结论（Schnotz & Baadte, 2015; Holyoak & Koh, 1987）。表面相似性对老年人的推理质量影响更大，而对于年轻人而言，结构相似性对其推理质量作用更大。**探讨表面相似性和结构相似性对老年人类比推理的作用是本研究关注的重点。**

**意见 2：**是否应该测试一下被试的工作记忆与智商？

**回应：**感谢您的建议，这一建议非常合理，本研究在进行时确实对被试基本认知能力进行测量：

- 1) 词汇理解。在韦氏成人智力量表的领悟力测验中抽取 10 个词语，让被试对其进行解释，计分方法参照韦氏成人智力量表，根据回答质量计 2、1 或 0 分，记录总得分，作为言语理解能力的指标。
- 2) 数字符号转换。取韦氏智力测验的数字符号转换分量表，被试需将每个数字对应的符号按顺序写下来，记录 90 秒内书写正确的个数，作为加工速度的指标。
- 3) 数字倒背。取自韦氏成人智力测验量表的数字倒背分量表，给被试念一串数字，让被试倒着背出来，先从数字个数较少的数字串开始，回答正确可依次加长，连续两次回答错误测验结束，记录正确倒背出的最长个数作为工作记忆广度的指标。
- 4) Stroop 测试。要求被试看打印出的彩色字，忽略字的读音并读出字的颜色。记录 90 秒内正确念出的颜色个数，作为抑制能力的指标。

老年人和年轻人都完成了所有认知能力测试任务，不同年龄组间的受教育年限和词语理解测验得分没有显著差异（ $p > 0.05$ ），可保证正确理解实验材料；年轻人在数字符号转换测验、数字倒背广度、Stroop 测验得分均好于老年人（ $p < 0.05$ ）。但我们的研究也发现人到老年，基本心理能力衰退是必然的，在年龄差异研究中，我们假设老年人与年轻人在高级认知过程中出现差异，背后机制或许是基本心理能力的衰退。

表 1 被试基本信息和背景测验汇总表（标准差）

	年轻人	老年人
年龄	21.86(1.432)	69.09(5.114)
受教育年限	14.48(1.022)	14.34(2.458)
词语理解得分	17.17(1.818)	16.65(2.496)
数字符号转换得分(个)	77.93(9.490)	44.97(9.687)
数字倒背广度(个)	8.24(1.327)	5.52(1.267)
Stroop 测验(个)	80.38(19.397)	48.78(17.270)

研究结束后，对这部分数据进行了分析，结果发现，在控制了这些基本认知能力后，推理成绩的年龄的主效应不显著[ $F(1,55) = 1.51, p = 0.23, \eta^2 = 0.027$ ]，表面相似性[ $F(1,55) = 1.08, p = 0.30, \eta^2 = 0.019$ ]和结构相似性[ $F(1,55) = 0.89, p = 0.35, \eta^2 = 0.016$ ]的主效应也不显著。我们反思，推理之所以出现年龄差异，就是因为老年人和年轻人在基本认知能力上有差异造成的，正是由于这些基本认知能力的差异带来的高级心理能力的衰退，如工作记忆和加工速度

(Morrison, 2005; Salthouse, 1992b,2010; 彭华茂, 2004)、抑制能力 (Conway, 2002) 等。

因此, 将基本认知能力作为协变量控制, 实际上是把主要差异来源控制掉了, 故在正式报告的数据分析结果中, 我们没有将基本认知能力作为协变量纳入。

Bugaiska, A., & Thibaut, J. (2015). Analogical reasoning and aging: The processing speed and inhibition hypothesis. *Aging, Neuropsychology, and Cognition*, 22(3), 340-356.

Conway, A. R., Cowan, N., & Bunting, M. F, et al. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163-183.

Morrison, R. G. (2005). Thinking in working memory, In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning*. Cambridge, England: Cambridge University Press.

Salthouse, T. A. (1992b). Working memory mediation of adult age differences in integrative reasoning. *Memory & Cognition*, 20, 413-423.

Salthouse, T. A. (2005). Effects of aging on reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning*. Cambridge, England: Cambridge University Press.

Salthouse, T. A. (2010). *Major Issues in Cognitive Aging*. New York: Oxford University Press.

彭华茂, 王大华. (2012). 基本心理能力老化的认知机制. *心理科学进展*, 20(8):1251-1258.

**意见 3:** 表 1 中的字母与数字如 F1 表示什么?

**回应:** 感谢您的建议, E = 实体 (entities); F = 初级关系 (first-order relation)

已在表格的注解中进行补充 (请见第 121 行, 方法部分 2.2 实验材料):

Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, 25(4), 524-575.

Catrambone, R. (2002). The effects of surface and structural feature matches on the access of story analogs. *Journal of Experimental psychology: learning, memory, and cognition*. 28(2), 318-334.

**意见 4:** 第 11 页中的“x 洒 y”是什么意思?

**回应:** 感谢您的建议, 这样写似乎太过抽象了, 在文中重新进行了解释: x 洒 y 代表实体间的主谓关系, 即初级关系, 如在引言第二段中的例子“老张情急之下赶紧把手中的汤洒在草上”, “x 洒 y”可以衍生出“老张洒水”, “李老头洒汤”等。这一说法源于 Catrambone (2002): *A first-order relation (FOR) match between stories would be one in which certain similar low-level predicates would be present in both stories (e.g., x SHOOT y; a FIRE-ON b)*. 在该研究中, *x SHOOT y* 可以从“猎人射鹰”衍生为“A 国攻击 B 国”

在正文中修改如下 (请见第 151 行, 方法部分 2.3 实验设计):



用谓语的匹配对数来量化结构相似性程度（如：x 着火，x 烤焦；x 泼 y，x 洒 y；Catrambone, 2002），如“老张洒汤”可以衍生出“老张洒水”，“李老头洒汤”等。

Catrambone, R. (2002). The effects of surface and structural feature matches on the access of story analogs. *Journal of Experimental psychology: learning, memory, and cognition*. 28(2), 318-334.

**意见 5：**学习阶段延迟三天后为测试阶段。为何要三天后？

**回应：**感谢您的提醒，已对研究设计的逻辑进行重新理清，并修改在文中相应位置。

**首先，在类比推理中，延迟时间会起到一定的作用。**首先，被试阅读大量的源故事，延迟一段时间后再阅读大量的靶故事，研究者感兴趣的问题正是源故事是否会对靶故事的解决起到提示作用；若有作用，则提示作用是由特定的源故事带来的吗，具体而言，对杯水车薪的故事解决是根据与之对应的源问题还是伯乐相马？更进一步，表面相似性和结构相似性的对靶问题解决的提示作用哪个更大些？即时的类比更多是在工作记忆范畴下的讨论，而延迟匹配则是长时记忆的作用，也是本研究更为关心的问题。

**其次，本研究在决定延迟天数时，**通过查阅文献，发现 Catrambone（2002）对年轻人类比迁移检索的研究，采用的延迟迁移间隔时间为 7 天，但是即使是相似度最高的源问题回忆率仅有 0.31，若本研究的延迟迁移间隔时间也设定为 7 天，可以预期老年人将会出现“地板效应”。因此，延迟间隔时间参照了张庆林和王永明（1999）的研究，设定为 3 天，结果中发现年轻人没有出现“天花板效应”，老年人也没有出现“地板效应”，因此，这一间隔时间是比较合适的。此外，间隔时间是 3 天，这就涉及到了长时记忆，控制变量的测量中发现即使长时记忆中再认词汇的成绩，老年人与年轻人没有区别，老年人的检索还是发生了严重衰退。所有理论都强调表面相似性对检索源问题的重要性，本研究的结果也证明了这一点：两个年龄组下，表面信息重叠水平的提高不仅能够促进表面信息的检索，也能促进结构信息的检索，表面信息的确对问题整体的提取有促进作用。

对于这一实验逻辑，已经补充在研究设计和结果部分，具体如下：

#### （请见第 159 行，方法部分 2.3 实验设计）

延迟三天后为测试阶段（Catrambone, 2002）。在类比推理中，即时类比更多是在工作记忆范畴下的讨论，而延迟匹配则是长时记忆的作用，故延迟时间会起到一定作用（Chen, Mo, & Honomichl, 2004）。被试在学习阶段阅读大量源故事，延迟一段时间后源故事是否会对靶问题的解决起到提示作用以及表面相似性和结构相似性对靶问题解决的提示作用哪个更大是本研究关注的问题。在决定延迟天数时，Catrambone（2002）的研究间隔时间为 7 天，但即使是相似度最高的源问题回忆率也仅有 0.31，可以预期老年人或会出现“地板效应”。本研究中延迟间隔时间设定为 3 天（张庆林，王永明，1999），结果发现年轻人没有出现“天花板效应”，老年人也无“地板效应”，故这一间隔时间较为合适。

#### （请见第 197 行，结果部分 3.1 推理质量的年龄特征）

学习与测验阶段间隔三天涉及到长时记忆，而两个年龄组在后测词语再认的正确率上没有显著差异 [ $t(59) = -0.548, p > 0.05$ ]，故检索源问题的分析若出现差异将排除是老年人

长时记忆较差的原因。

Catrambone, R. (2002). The effects of surface and structural feature matches on the access of story analogs. *Journal of Experimental psychology: learning, memory, and cognition*. 28(2), 318-334.

Chen, Z., Mo, L., & Honomichl, R. (2004). Having the memory of an elephant: Long-term retrieval and the use of analogues in problem solving. *Journal of Experimental Psychology: General*, 133(3), 415-433.

张庆林, 王永明. (1999). 类比迁移发生机制的研究. *心理科学*, 22(2), 141-143.

**意见 6:** 如被试 A 在测试阶段, 完成杯水车薪 E3F3、伯乐相马 E1F3、恰到好处 E3F1、叶公好龙 E1F1; 被试 B 完成杯水车薪 E1F1、伯乐相马 E3F3、恰到好处 E1F3、叶公好龙 E3F1; 被试 C 完成杯水车薪 E3F1、伯乐相马 E1F1、恰到好处 E3F3、叶公好龙 E1F3; 被试 D 完成杯水车薪 E1F3、伯乐相马 E3F1、恰到好处 E1F1、叶公好龙 E3F3。这里似乎不清楚, 是被试间设计? 即不同被试完成不同类型的推理问题。

**回应:** 研究为拉丁方设计, 也即轮换设计。表面相似和结构相似均为被试内变量, 故而同一问题情境设置有四种不同的相似性条件 (E3F3 E3F1 E1F3 E1F1), 但不能让被试对同一情境的问题进行 E1F1~E3F3 的 4 种相似性条件下的反复推理, 所以只能采用拉丁方设计, 让被试接受全部实验处理, 但不是重复接受同样的材料, 也即相似的问题情境。

被试接受靶问题操纵分布情况如下表:

故事内容	故事 1	故事 2	故事 3	故事 4
被试分区	(杯水车薪)	(伯乐相马)	(恰到好处)	(叶公好龙)
Block1	E3F3	E1F3	E3F1	E1F1
Block2	E1F1	E3F3	E1F3	E3F1
Block3	E3F1	E1F1	E3F3	E1F3
Block4	E1F3	E3F1	E1F1	E3F3

已在文中“方法”部分进行阐释 (请见第 167 行, 方法部分 2.3 实验设计):

每个源问题都对应 4 种相似水平不同的靶问题, 为避免被试重复回答类似情境问题且对同一情境的问题进行 4 种相似性条件下的反复推理, 故采用拉丁方设计: 在测试阶段, 每个被试看到 4 个源问题各自对应的 4 个不同相似水平的靶问题中的 1 个, 并保证每个被试最后均接受 4 种相似水平的不同靶问题。如被试 A 在测试阶段, 完成杯水车薪 E3F3、伯乐相马 E1F3、恰到好处 E3F1、叶公好龙 E1F1; 被试 B 完成杯水车薪 E1F1、伯乐相马 E3F3、恰到好处 E1F3、叶公好龙 E3F1; 被试 C 完成杯水车薪 E3F1、伯乐相马 E1F1、恰到好处 E3F3、叶公好龙 E1F3; 被试 D 完成杯水车薪 E1F3、伯乐相马 E3F1、恰到好处 E1F1、叶公好龙 E3F3。总的来说, 被试接受的实验处理是一样的, 只是具体处理对应的故事情境不同。

**意见 7:** 16 个填充材料具体是什么? 为何要设置这些材料?

**回应:** 填充材料来自于 Gentner et al. (1993), 经过修改后作为填充材料。每个源问题和靶问题之间共享一个高级结构关系 (HOR) 的匹配。研究范式源自 Catrambone (2002), 在他的

研究中填充材料和实验材料的比例也是 1：5。

设置填充材料的**目的**在于：

- 1 防止被试直接关联源问题和靶问题。被试阅读大量的源故事，延迟一段时间后再阅读大量的靶故事，研究者感兴趣的问题正是源故事是否会对靶故事的解决起到提示作用；若有作用，则提示作用是由特定的源故事带来的吗，具体而言，对杯水车薪的故事解决是根据与之对应的源问题还是伯乐相马？
- 2 使得研究更具有生态效度。在现实生活中遇到不同情境的问题时，需从一堆先前遇到过的问题情境中发现可类比的情境，但往往只有一部分问题是拥有先前解决经验，知道解决办法的。故而为了模拟生活经验，设置了诸多填充材料。

在正文中补充在 2.2 实验材料部分，如下（请见第 127 行，方法部分 2.2 实验材料）：

填充材料来自于 Gentner 等 (1993)，经修改后作为填充材料。每个源问题和靶问题之间共享一个高级结构关系（HOR）的匹配，填充材料和实验材料的比例是 4:1（Catrambone, 2002）。设置填充材料的目的在于：（1）防止被试直接关联源问题和靶问题。被试阅读大量的源故事，延迟一段时间后再阅读大量的靶故事，研究者关注源故事是否会对靶故事的解决有提示作用；若有作用，则提示作用是由特定的源故事带来的吗？（2）使研究更具生态效度。在现实生活中遇到不同情境的问题时，需从先前遇到过的问题情境中发现可类比的情境，但往往只有一部分问题是拥有先前解决经验的，为模拟生活经验，设置填充材料。

Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, 25(4), 524-575.

此外，研究中使用到的所有填充材料见下图（word 版已作为附件重新上传）

源问题	靶问题	HOR
小狗宾利嘴里叼着一块肉，正沿着码头跑着。但是，一不小心，肉掉进湖里了，它没有晚餐可以吃了。宾利看到另一只狗也叼着一块肉，就想接近那条狗，然后把肉抢过来。不幸的是，这只狗块头比它大多了。最后，宾利在打斗中失去了牙齿。	查理戴着一块昂贵的手表，经过一家旅馆。当他在行走的途中，手表不小心滑到了下水道里，这下他没有手表可带了。查理看到另一个人也带着昂贵的手表，就想接近那人，把表偷过来。糟糕的是，那人恰巧是个警察。最后，_____。	“不自量力”
威廉是一个精神病院里的病人，几乎整天都呆在室内，把所有时间都花在关于食物的白日梦上。每个月的例行房间卫生检查他都无法通过，所以他很讨厌医生和护士。在四月例行检查的前几天，他的房间还是一团糟，为了鼓励他整理，护士答应如果他清洁并整理好房间就给他买饼干吃。听到这个消息后，威廉非常高兴。但是距离检查的时间太短了，而且威廉还是整天做白日梦。最终，他还是没有通过检查，也没有得到饼干。	卡伦是个中学生，她不好好学习，成绩很差，还鄙视认真学习的人。她非常喜欢假期，并且整天幻想着去夏威夷玩。高一期末考试前几周，她还什么都没学，因为天天想着玩。她爸爸为了激励她学习，答应她如果考前好好学习，并且期末考试达到一定的成绩，就出钱让她去夏威夷玩。这让卡伦非常开心，但是她的功课落下的实在太多，而且这几个星期她一直在幻想着去夏威夷要怎么玩。最后，_____。	“白日做梦，一事无成”
牛先生是一家剃须刀公司的经理。他们公司有发明家，将刀片的材质进行完善，使其比以往的任何剃刀都锋利。然而，剃刀材质改进后，由于太锋利，很容易将使用者割伤。有一次，牛先生在使用的时候将自己严重割伤了，还好没有感染。这件事情让发明家意识到需要开发剃刀的安全功能，避免使用者伤到自己。牛先生震惊于剃须刀的杀伤力，也认为安全功能是非常必要的。	博伊斯女士在一家日光灯公司负责销售。1980 年，他们公司的一名供应商将灯丝的材料进行完善，使其比以往的灯丝发出更亮的光。由于会使灯变得很烫，供应商贴出警告标签，避免使用者伤到自己。博伊斯女士认为不用贴标签，因为会增加工作量。有一次，她在用这些新的灯的时候不小心烫伤了。她_____。	“吃一堑长一智”（悔不当初）
马先生是一个亿万富翁，他雇佣了一名司机开着他的豪车接送他。他经常对他的妻子说他不会迟到的，因为他有司机。有一天早上马先生急急忙忙的要赶去参加一个会议，司机却睡着了，因为他以为这天不用接送。马先生迟到了很久，为了避免类似情况再发生，他决定雇佣第二名司机。这要花一大笔钱，因为第二名司机可能要更高的工资。	很久以前，有个地保和他的太太住在一起，妻子负责做饭。他每次都和朋友吹嘘说有太太，在家里吃饭都可以吃得很好。有一天，他非常饿的回到家里准备吃饭，但是他太太以为他会外面吃，就没有准备他的饭菜。这个地保没吃晚饭就出门了，为了避免以后再出现这种情况的挨饿，他决定娶第二房太太。花费增加了很多，_____。	“得不偿失”

一只老鹰住在高大的橡树上。有一天它看到地面上有个猎人，用没有羽毛的箭瞄准它，但是没射中。老鹰知道羽毛能提高箭的命中率，于是它到树下给了猎人一些羽毛。果然，猎人的箭附上羽毛后命中率提高了不少。猎人很感激老鹰，承诺再也不猎杀它。以后，猎人转而以猎杀鹿为生。	曾经有个小国——Z 国，有着世界上最先进计算机技术。有一天它被强大的邻国——G 国突然袭击，但是由于导弹定位技术很差，所以袭击失败了。Z 国的政府意识到，他们国家先进的计算机技术可以提高导弹定位技术，于是他们将计算机技术引进到 G 国。G 国的导弹定位经过改进后，提高了很多。G 国非常满意，并承诺不再袭击 Z 国。以后，	“知恩图报”
彼得是一个来自冰岛（寒带）的年轻人，某个夏天他搬到了美国的佛罗里达（热带）。他对自己的肤色非常在意，他认为当地人可能不会接受他，除非他把自己晒得和其他人一样黑。所以，彼得花了一整天晒太阳。但是他不了解阳光的毒辣，到晚上他的皮肤就中度灼伤了，不得不去医院治疗。彼得认为佛罗里达不适合他，并决定预订下一班船返回冰岛生活。在佛罗里达剩下的时间里，他都打着一把大阳伞防晒。	Q 公司是一个生产电器元件的小公司，有一年他们决定将公司转型去生产集成电路。Q 公司认为要提高竞争力，就必须拥有复杂机械、技术的生产车间。于是 Q 公司将所有的经费都投入到车间改造上面。但是他们没有意识到这一举动带来的资金风险，六个月内，公司亏损得非常严重。该公司认为他们不适合生产集成电路，还是慢慢恢复成原来的车间，继续生产电器元件。在离车间完全恢复的一段时间内，	“欲速则不达”
一头骡子发现窗外有几个梨子。它心想：“这些梨子似乎烂掉了，也许我可以摘几个下来，看看我的预测是否正确。”但是，那些梨子对他来说太高了，并且他实在饿得很虚弱，跳起来也没法够着梨子。当然，骡子非常失望。最后它闭上眼睛睡着了，梦想着拥有山一样多的梨子。	一个女孩在音像店发现了一些令人好奇的 CD。她暗自道：“这些 CD 看起来很糟糕，我可以把它们买下来，看看我的预感是否正确。”但是，那些 CD 对她来说太贵了，思前想后，她也没有其他办法筹到钱把这些 CD 买下来。不用说，她感到相当失望。最后，	“痴心妄想”
天气温暖的时候，知更鸟整天都在喳喳叫。天气渐渐冷了，知更鸟拜访了松鼠，唱了一首歌给松鼠听，希望得到一些葵花籽作为回报。然而松鼠听完以后表示非常失望。它对知更鸟喊道“你的歌声太可怕了，我是不会给你任何葵花籽的！”知更鸟哭了，发誓以后再也不唱歌了。	山姆有钱的时候就跑去环游世界，买了很多漂亮的东西。当他的钱快花完的时候，他拜访了他妈妈，并送他上在西藏旅游的时候买的礼物，希望能够从他妈妈那拿点钱。但是，他妈妈一点都不满意。她对山姆说“你给我买的是什么垃圾，你从我这一分钱也拿不到！”山姆哭了，	“偷鸡不成蚀把米”
小狗宾利嘴里叼着一块肉，正沿着码头跑着。但是，一不小心，肉掉进湖里了，它没有晚餐可以吃了。宾利看到另一只狗也叼着一块肉，就想着接近那条狗，然后把肉抢过来。不幸的是，这只狗块头比它大多了。最后，宾利在打斗中失去了牙齿。	查理带着一块昂贵的手表，经过一家旅馆。当他在行走的途中，手表不小心滑到了下水道里，这下他没有手表可带了。查理看到另一个人也带着昂贵的手表，就想着接近那人，把表抢过来。糟糕的是，那人恰巧是个警察。最后，	“不自量力”
鲍勃和伊万都认为对方的专业技能很好，于是决定共同开一家专卖店。由于运气不太好，再加上伊万的疏忽，他们损失了一大笔钱。鲍勃很生气，让伊万不要再插手专卖店资金的事情。后来，伊万将他开办的一个装置卖掉了，填补了之前由于他的疏忽而造成的资金漏洞。	一对男女互相相爱，都认为对方是对的人，于是决定结婚，组成自己的小家庭。婚后，由于路况不好，外加妻子的莽撞，把丈夫的新车撞出了一个深坑。这让丈夫非常的生气和郁闷，再也不要让妻子开他的车了。后来，妻子	“将功补过”
有两个小国家——B 国和 C 国，他们有一个强大且好战的邻国——M 国。B 国一直在思考对自己最有利的局势，想要利用并吞并 C 国，而 C 国还来不及寻求其他强国的保护就被 B 国控制了。胜利的 B 国向强大且好战的邻居——M 国提出一个协议，他们把 C 国献给 M 国，但 M 国要让 B 国维持独立。但是 M 国的反应大大超乎 B、C 国的预期，而 B 国忙于镇压 C 国，完全没有精力阻止 M 国的举动。最终，M 国吞并了这两个小国家，并设立了傀儡政府控制舆论。	两个初中生——林肯和莫兰，一起去上学，在路上遇到了一个高中生——扎德，他很喜欢欺负初中生。莫兰一直在思考如何利用林肯，来达到对自己最有利的局面。林肯想跑去找警察，然而才刚跑出去就被莫兰抓回来。莫兰把林肯推到扎德面前，让扎德欺负林肯就好，放过自己。扎德的反应大大超乎莫兰的想象，莫兰一直在防止林肯逃跑，一点自救的措施都没做，根本招架不住扎德。最后，	“唇亡齿寒”
一个农夫在他果园的其他空地上种上了他喜欢的杏树，每年他都将果实作为礼物送给他认识的人。但是有一年，树都枯萎了，颗粒无收，所有人都很不开心。农夫认识的一个法官，特别喜欢吃这些杏树结出的果实，他决定把这些树救活。于是，他付给农夫一大笔钱，将这些杏树移植到自己的地里。换了土壤后，这些树都活过来了，又再次结出了好吃的果实。第二年，这些杏树结出的果实参加农产品大赛获得了第一名。	一个绅士在自己的宠物里又增加了他喜欢的金丝雀，隔三差五的就邀请朋友过来听金丝雀唱歌。然而一个秋天，金丝雀的情况变得有些糟糕，不能唱出好听的歌，所有人都都很郁闷。绅士有个兽医朋友，非常喜欢金丝雀的歌，他决定治好金丝雀。他付给这个绅士很多钱，买下金丝雀，并带回家亲自照顾。换了环境后，金丝雀又能够唱出美妙的歌了。第二年，	“塞翁失马焉知非福”
从前有个头脑简单的囚犯，特别喜欢晚上在牢房里吹口琴。终于有一天，他的狱友们厌倦了他每天夜吹口琴，让他赶紧闭嘴，要吹就白天的时候天天吹。第二天，这名囚犯越狱了，并躲进了一辆篷车里。这时他突然想起他的狱友让他“天天吹口琴”，于是拿出口琴开始吹。结果他立即就被抓住了。这个头脑简单的囚犯再也不想像睬他的狱友，于是天天在牢房墙壁上涂鸦。	珍妮和她的狗每天在邻居的院子里玩。终于，她的邻居受够了，告诉珍妮要天天玩可以，但应该在她自己家的院子里天天玩。过了几天，珍妮没做家务就偷偷溜出家门，躲在操场上玩。但是她想起来邻居说她应该天天玩，于是她跑回家。结果她立马被她妈妈抓住了。这个小女孩	“断章取义”
牧羊人在森林边缘驱赶他的羊，突然有只狮子从森林中窜出来。牧羊人吓了一跳，开始朝狮子大吼，希望能把狮子吓走。但是大吼大叫不仅没把狮子吓走，反而吸引了狮子的注意力。狮子向他走来，他才意识到刚才应该不要理会狮子的。为了逃跑，牧羊人扔给狮子一只绵羊，趁着狮子吃羊的功夫，他带着其余的羊群赶快离开，直到他觉得离狮子已经足够远了才停下来。他认为和狮子的距离已经在安全范围内了，于是他在草地上打了个盹。当他醒来的时候发现所有的羊都不见了，到处找都找不到。这件事过后，牧羊人决定去买一只大口径的猎枪，下次放羊的时候随身带着。	天文学家非常痴迷于观察星象，渐渐地，他不断的改建自己的房子，想把它变成一个天文观测台。他的妻子抱怨说他把太多的时间都花费在改造房子上，而忽略了他自己的本职工作——天文学。这天天文学家觉得妻子说的很正确，于是停止改造房子，整天只思考天文学问题。有天晚上，他穿过他的房子的时候，脚下有块还未完工的天花板塌了，他受伤严重。事故让他非常后怕，也认识到思考天文学的同时也必须兼顾改造房子。最后，他	“亡羊补牢”
有个皇帝非常沉迷于战争和敛财。有一年，他决定重建国中所有的桥梁，让他们足以承载他所有的军队。大臣们抱怨说皇帝花了太多的钱在建桥梁上，却没有足够的钱发动战争。皇帝觉得很有道理，于是放弃改建桥梁的计划，决定全面侵略邻国。有一天，皇帝坐着马车经过一段还未完全建成的桥梁时，马车坠毁了，他本人受了很严重的伤。这让他心有余悸，并清楚的认识到必须平衡建桥梁支出费用和军费的平衡。最后，他还让他的臣子重新打造一项皇冠，因为旧的那项在坠毁时摔坏了。	守财奴在他家的后花园点钱，突然发现有个乞丐正接近花园的灌木丛。守财奴非常惊慌，所以大声的呵斥那个乞丐赶紧离开。但是大声呵斥不仅没把乞丐赶走，反而吸引了乞丐的注意力。当那个乞丐开始朝守财奴走来，他才意识到刚才应该不要管那个乞丐的。为了赶紧逃离现在这个处境，守财奴扔了一个硬币给那个乞丐，趁着乞丐捡钱的功夫，他带着自己的那袋钱逃回了房子里，并锁上门。守财奴觉得在房子里应该够安全了，于是他去睡了个午觉。当他醒来的时候，发现他的那袋钱不见了，怎么找都找不着。这件事过后，	“得寸进尺”
一个专门养鸡的农场主正站在鸡舍里捡鸡蛋，所有的鸡都在鸡舍外觅食。当这些鸡觅食回来后，不知道什么原因都开始攻击这个农场主。农场主生气的大喊“你们这群忘恩负义的畜生！”但是所有的鸡都在驱逐他，一直把他赶到了树丛边上，农场主很受伤。一个小偷趁着群鸡驱赶农场主的时候，溜进鸡舍，把鸡蛋都偷走了。当农场主从医院治疗结束回来后，他把养鸡场卖去非洲度假了。	一个夏令营的教官，想趁着营员去爬山的时候，去附近的沙滩探查一下露营的可行性。当他站在沙滩上的时候，营员都回来了，开始缠着他问“晚饭吃什么？”“可否去游泳？”等问题。教官大喊“你们为什么总在我想要认真考虑我们可以做什么的时候来纠缠我？”他风一样的跑离海滩，但是营员们还是在后面追着他，嚷嚷道“有点耐心嘛！”一个大浪打来，把沙滩上的睡袋都卷走了。带完这一期夏令营后，教官	“忘恩负义”
有两个冒险者 A 和 B，他们同时发现了一个无人居住的漂亮的小山谷，都想占为己有。他们俩为了谁是第一个发现这个山谷争论不休，甚至大打出手。当他们正在争执的时候，一个很有威信的仲裁者出现了，将山谷划分为两段，一人一半。A 和 B 意识到他们早该这么划分。从那时起他们俩就不再争执，还成为很好的朋友，在山谷中互相帮助对方建房子。	迎面相向走来两个人，同时发现了地上有一张 100 元钞票，都想把这钱占为己有。他们俩对谁先发现钱这件事各执己见，互不相让。当他们正在争执的时候，警察来了，捡起 100 元，换出两张 50 元的分给他们，一人一张。这两个人才意识到他们可以这样把钱分了。后来，	“双赢”

**意见 8：**文中有多处词语重复或标点符号的问题。如第 13 页第 1 行。

**回应：**感谢您的建议，已对全文进行阅读，并请类似研究领域的专家进行了批判性阅读，对已有文法错误进行修改。

**意见 9：**图 1 的标题不能很准确地概括图的内容。

**回应：**您的建议非常合理，已修改为：**年轻人-老年人在不同相似性条件下的推理质量（请见第 218 行，结果部分 3.1 推理质量的年龄特征）**

**意见 10：**作者在讨论部分指出“由此可见，在使用更具生态效度的问题解决范式时，老年人的推理成绩有所提升，与年轻人差距更小。”“不同研究之间的比较的可比性值得商榷。

**回应：**感谢您的建议，总的来说，这种比较方法源于追踪研究中对不同同辈群体的比较，如：在 BLSA（Baltimore Longitudinal Study of Aging, BLSA）追踪样本中，研究者对不同同辈群体之间的幸福感进行了比较，使用了标准差的形式，发现 1935 年的同辈群体与 1925 年同辈群体间相差了 0.20 个 SD（Hülür, Ram, & Gerstorf, 2015）；在 Schilling（2005, Sample B）中发现 1919 年的同辈群体比 1909 年同辈群体低了 0.16 个 SD 等。

追踪研究中不同同辈群体的比较可行，在于使用了相同的研究工具；本文的讨论部分主要想比较两种研究范式的不同。首先，我们对被试的同质性进行分析，发现本研究的被试与文章进行对比的两组被试之间存在一定差异：

	本研究		高悦等，2014		Salthouse，1992b	
	年轻人 (N=29)	老年人 (N=32)	年轻人 (N=74)	老年人 (N=43)	年轻人 (N=30)	老年人 (N=30)
年龄	21.86±1.43	69.1±5.11	21.62±2.30	84.53±4.92	20.1±1.0	68.2±5.7
受教育年限	14.48±1.02	14.34±2.46	15.09±1.74	6.90±4.25	14.2±1.1	14.5±2.2
词语理解	17.17±1.82	16.65±2.50	-	-	-	-
数字符号转换	77.93±9.50	44.97±9.69	-	-	-	-
数字倒背广度	8.24±1.33	5.52±1.27	7.26±1.69	3.40±1.43	-	-
Stroop 测验	80.38±19.40	48.78±17.28	-	-	-	-
推理得分	0.62±0.28	0.38±0.25	6.82±1.79	1.08±0.87	79.1±11.1	59.9±9.2
差异系数 CV	45%	66%	26%	81%	14%	15%

由上表可知，本研究在取样上，与 Salthouse（1992b）的样本更为相似（年龄和受教育水平基本相当）。差异系数由于是相对差异量数，它既可用于不同单位资料的差异比较，也可用于不同水平的同类现象的差异情况的比较，由上表可看出高悦等（2014）研究中的老年人样本离散性较大，与年轻人被试相差较多；而 Salthouse（1992b）的研究中两年龄组的被试的离散性相差较小。本研究的离散性处于两者之间。

为检验两种范式的差异，我们进行了数据模拟。首先，使用 Matlab R2012b 生成了 6 列与三项研究具有相同样本量、均值和标准差的随机数据；以实验处理方式和年龄组为被试间变量分别进行 2 实验处理（经典推理范式/问题解决范式）×2 年龄组（年轻人/老年人）方差分析。



结果发现，年龄组与实验处理交互作用均显著。

#### ■ 本研究与高悦等（2014）的研究对比

年龄组与实验处理交互作用显著 $[F(1,177) = 158.35, p < 0.001, \eta^2 = 0.476]$ 。简单效应分析发现：在经典推理范式中，年轻人与老年人的得分具有显著差异 $[F(1,174) = 516.80, p < 0.001, \eta^2 = 0.748, MD = 5.925]$ ；在问题解决范式中，年轻人与老年人的得分没有显著差异 $[F(1,174) = 1.66, p = 0.199, \eta^2 = 0.009, MD = 0.449]$ 。

#### ■ 本研究与 Salthouse（1992b）的研究对比

年龄组与实验处理交互作用显著 $[F(1,120) = 107.37, p < 0.001, \eta^2 = 0.479]$ 。简单效应分析发现：在经典推理范式中，年轻人与老年人的得分具有显著差异 $[F(1,117) = 228.81, p < 0.001, \eta^2 = 0.655, MD = 23.011]$ ；在问题解决范式中，年轻人与老年人的得分没有显著差异 $[F(1,117) = 0.09, p = 0.770, \eta^2 = 0.001, MD = 0.449]$ 。

总之，从结果来看，问题解决范式中，老年人和年轻人推理成绩差异相比经典推理范式显著变小，这一模拟过程可从侧面进一步说明这种研究范式有助于老年人问题解决，或至少缩短了与年轻人之间的差异。非常感谢您的建议，确实跨研究数据之间的比较是不够严谨的，只能作为粗略的探讨。方法的局限性已在正文及参考文献中进行补充

（请见第 248 行，讨论部分）：

首先，类比推理出现了老化，与前人研究一致（Aichelburg et al., 2016; Bugaiska & Thibaut, 2015; Salthouse, 2005），但本研究发现的基于问题解决式的类比推理成绩的年龄差异，相比于经典类比推理任务的年龄差异较小。对比两种范式，以年轻人推理成绩的标准差为标准，用两年龄组的各自推理得分之差除以该范式下年轻人标准差（Gerstorff, Ram, Hoppmann, Willis, & Schaie, 2011），以此衡量不同任务范式下推理成绩的年龄差异。可以发现，在经典类比推理范式下，老年人的成绩比年轻人低 3.20SD（老年人：1.08±0.87；年轻人：6.82±1.79；高悦，彭华茂，王大华和文静，2014）或 1.73SD（老年人：59.9±9.2；年轻人：79.1±11.1；Salthouse, 1992b）；而在本研究中老年人的成绩比年轻人低 0.86SD（老年人：0.38±0.25；年轻人：0.62±0.28）。由此可见，在使用更具生态效度的问题解决范式时，老年人的推理成绩有所提升，与年轻人差距更小。总的来说，这种比较方法源于追踪研究中对不同同辈群体（cohort）的比较（Hülür, Ram, & Gerstorff, 2015; Schilling, 2005），但跨研究数据之间的比较是不够严谨的，只能作为粗略探讨。

Gerstorff, D., Ram, N., Hoppmann, C., Willis, S. L., & Schaie, K. W. (2011). Cohort differences in cognitive aging and terminal decline in the seattle longitudinal study. *Developmental Psychology*, 47(4), 1026-1041.

Hülür, G., Ram, N., & Gerstorff, D. (2015). Historical improvements in well-being do not hold in late life: Birth-and death-year cohorts in the United States and Germany. *Developmental psychology*, 51(7), 998-1012.

Schilling, O. K. (2005). Cohort- and age-related decline in elder's life satisfaction: is there really a paradox? *European Journal of Ageing*, 2(4), 254-263.

**意见 11：**英文摘要有多处错误。

**回应：**谢谢您，抱歉出现了这么多错误，已经请两位北京语言大学的语言学硕士帮忙修改，并请英语母语者帮忙校对。

---

## 第二轮

**审稿人 3 意见：**

**意见：**作者认真地回答了我在一审时所提问题，可考虑录用。

**回应：**非常感谢您对本研究提出的宝贵意见，这使得我们的文章在严谨性和科学性上有了很大的提升。感谢您的首肯，祝您研究愉快！

**编委复审意见：**

**意见：**建议发表。请作者结合各位审稿人的历次修改意见，再次认真审读，以最大程度减少失误、提高稿件质量。

**回应：**非常感谢您对本文的首肯，已对全文进行批判性阅读，同时按照《心理学报》的要求修改了标点符号、参考文献格式等。祝您研究愉快！